Technische Universität Braunschweig
Institut für Informationssysteme
http://www.ifis.cs.tu-bs.de

Christoph Lofi

Distributed Data Management

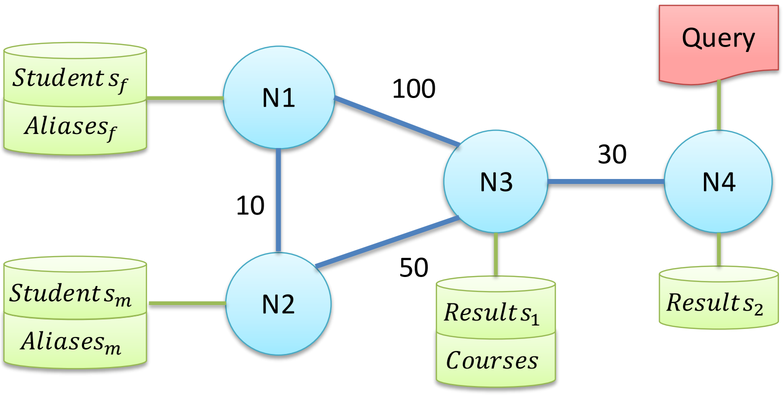
Sheet 3

Exercise 1

Let's consider a beloved database schema from the RDB1 lecture… (see Appendix A).

1. Provide an SQL query to return all course titles and respective exam results of Mystique
2. Transform the query into an operator tree and perform some simple algebraic optimizations (use heuristics, no distribution and cost models in this sub-exercise).
3. Now, let us assume that Students is horizontally fragmented by sex, Results vertically fragmented into two fragments (actually, this will result into one fragment containing only the primary key attributes and another fragment additionally containing the results; thus this fragmentation is more a partial replication…). Aliases uses horizontal partitioning derived from Students. Courses is not partitioned.
   Provide a **generic algebra statement** and a r**educed algebra statement** for the previous query (no graph – just algebra)**.**
4. Let's assume there are 4 nodes.
   - Node 1 allocates female $Students_f$ and $Aliases_f$.
   - Node 2 allocates male $Students_m$ and $Aliases_m$.
   - Node 3 allocates $Results_1$ (only keys) and $Courses$
   - Node 4 allocates $Results_2$ (keys and results)
   - Also, our query from exercise 1) originates from node 4.

   Find an optimized query plan with respect to overall time units used in the whole system.
   In this plan, mark clearly **at which node** the operations are performed and indicate data transfer with send-receive operations (example see slide 32 lecture 3).
   How **expensive** is your plan based on the provided **statistics**? How expensive is your plan in when executed on the **data** provided in the tables of the appendix?
   How much time is spent on **communication**? How much time is spent on **computation**?



Assume the following:

- You may use any **hybrid shipping** scheme you like
- Shipping queries or query fragments is free.
- The **communication costs** in time units (TU) for sending one tuple (size does not matter) between two nodes can be found in the above diagram.
    - There is no benefit from tuple blocking in communication
- Nodes may also **re-route tuples** (receive-send without any other operations).
    - No direct communication between N4 and N1 & N2
- **Selections** and **Projections** are pipelined and cost 1 TU each.
- **Accessing** (i.e. iterating over) relations is free, there are no indexes.
- **Joins** are performed using the block nested loop algorithm (i.e.: no pipelining, costs for $A \bowtie B$ are $|A| * |B|$ TUs).
- Only following **statistics** are known: (no additional information on the actual database content is known to the optimizer! Using additional information from Appendix A is cheating – no chocolate for that…!)
    - Number of tuples of each **relation**
    - Direct selections on the global relation $Students$ or $Aliases$ will return only one tuple (full names, matrikelnumbers, and aliases are unique)
    - Sex is either male or female. It is assumed that the ratio between both is 1:1.
    - Usually, students have one alias, but may be more (which is very rare).
    - In average, each student takes two exams

5. What is the first-tuple response time of your plan? No guessing – please show clearly how you computed the response time

Appendix A

Students

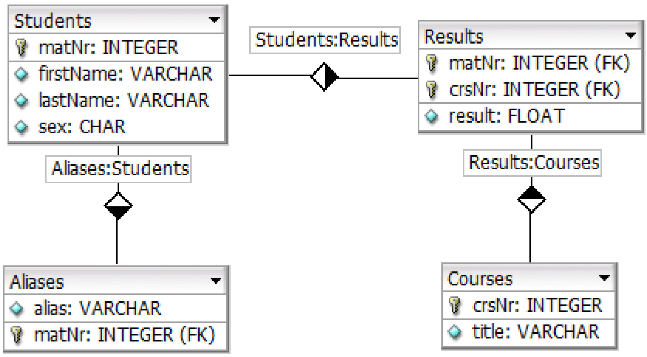
matNr	firstName	lastName	sex
1000	Clark Joseph	Kent	m
1001	Louise	Lane	f
1002	Lex	Luthor	m
1003	Charles	Xavier	m
1004	Erik	Magnus	m
1005	Jeanne	Gray	f
1006	Ororo	Munroe	f
1007	Tony Edward	Stark	m
1008	Matt	Murdock	m
1009	Raven	Wagner	f
1010	Robert Bruce	Banner	m

Results

matNr	crsNr	result
1009	100	3.7
1002	102	5.0
1000	101	4.0
1000	100	1.3
1004	102	1.3
1003	101	1.7
1007	103	3.0
1006	100	1.7
1009	102	1.3
1003	103	1.0
1009	101	1.0
1008	101	1.7

Aliases

alias	matNr
Mystique	1009
Daredevil	1008
Kal-El	1000
Professor X	1003
Hulk	1010
Windrider	1006
Superman	1000
Phoenix	1005
Ironman	1007
Magneto	1004
Mockingbird	1002
Storm	1006
Golden Avenger	1007
Queen of Wakanda	1006



Courses

crsNr	title
100	Introduction to Superheroism
101	Secret Identities 2
102	How to take over the world
103	Codes of Justice